

Generalized Γ calculus and application to interacting particles on a graph.

Pierre Monmarché

October 22, 2015

Abstract

The classical Bakry-Émery calculus is extended to study, for degenerated (non-elliptic, non-reversible, or non-diffusive) Markov processes, questions such as hypoellipticity, hypocoercivity, functional inequalities or Wasserstein contraction. In particular we obtain the optimal speed of convergence to equilibrium for any ergodic Ornstein-Uhlenbeck process, which is given by the spectral gap of the drift matrix and the size of the corresponding Jordan blocks. We also study chains of N interacting overdamped particles and establish for their invariant measures log-Sobolev inequalities with constants of order N^2 , which is optimal.

1 Introduction and overview

The Γ calculus has been introduced by Bakry and Émery in [4]. When L is a Markov generator, define the carré du champ operator by

$$\Gamma(f, g) = \frac{1}{2} (L(fg) - gLf - fLg)$$

and the Γ_2 operator by

$$\Gamma_2(f, g) = \frac{1}{2} (L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(g, Lf)).$$

These operators are related to various properties of L , of its associated semigroup $(P_t)_{t \geq 0} = (e^{tL})_{t \geq 0}$ and, as the case may be, its invariant measure μ : hypercontractivity, regularity, long-time behaviour, concentration of measure, etc. The main reference on this topic is certainly the book [5] of Bakry, Gentil and Ledoux; nevertheless in the rest of this work we will mostly refer to [9], which allow for a shorter introduction. This classical approach is for instance very efficient to study the reversible elliptic diffusion with generator

$$L = -\nabla U \cdot \nabla + \Delta$$

where Δ is the Laplace-Beltrami operator of a Riemannian manifold \mathcal{M} and U is a smooth potential on \mathcal{M} ; especially when U is strongly convex.

While this theory is still a vivid area of research, in parallel of its developments, interest has grown for some decades in so-called "degenerated" processes for which the classical methods do not work, at least not in their usual form. The typical example is the kinetic diffusion on \mathbb{R}^{2d} with generator

$$Lf(x, y) = [y \cdot \nabla_x + (\nabla_x U(x) + y) \cdot \nabla_y + \Delta_y] f(x, y),$$

which has inspired an extensive amount of works, among which we will only cite for now the memoir [30]. Piecewise deterministic Markov processes form another large class of such degenerated processes (see [24]).

The highlights of Villani's memoir include the idea that when classical quantities (variance, entropy, energy...) do not behave well under the action of the semigroup, it may be helpful to work with ad hoc distorted quantities for which usual techniques may apply. This heuristic have proved useful in a variety of situations, in various ways (see for instance [2, 15, 26]). The drawback is a (possibly drastic) complexification of the computations.

The aim of this paper is to give a somewhat unified framework in which the computations appear relatively nice. This is done in Section 2. The strategy closely follows the Bakry-Émery theory (and therefore we call it "generalised Γ calculus"), except we consider a broader variety of functionals. This idea is already present in [7, 2, 26].

For the sake of clarity we will try to avoid technicalities, and in particular we won't address the question of the optimal functional spaces for which our results hold. All the test functions will be considered in a space \mathcal{A} for which very strong assumptions will hold. In applications, it may often be taken as the set of \mathcal{C}^∞ functions which are bounded below by a positive constant and whose derivatives all grow at most polynomially at infinity.

In order to give an overview of possible methods, examples of applications will systematically be provided. Most of them come from existing works. As we will see the Γ calculus point of view sometimes improves and almost always simplifies them. For instance, consider the case of Ornstein-Uhlenbeck processes, with generator

$$Lf(x) = -(Bx) \cdot \nabla f(x) + \operatorname{div}(D \nabla f)(x),$$

where B and D are constant matrices. Suppose $\ker D$ does not contain any non trivial subspace which is invariant by B^T , and $\rho = \inf\{\Re(\nu), \nu \text{ eigenvalue of } B\} > 0$. Then the associated semigroup $(P_t)_{t \geq 0}$ admits a unique invariant measure μ . Denote by

$$\operatorname{Ent}_\mu f = \int f \ln f d\mu - \int f d\mu \ln \int f d\mu$$

the relative entropy of a function f with respect to μ . Arnold and his co-authors ([2, 1]) proved (working with L' the dual of L with respect to the Lebesgue measure, but this boils down to the same) that for all $\varepsilon > 0$ there exists a constant $c_\varepsilon > 0$ such that for all f with $\operatorname{Ent}_\mu f < \infty$,

$$\operatorname{Ent}_\mu(P_t f) \leq c_\varepsilon e^{-2(\rho-\varepsilon)t} \operatorname{Ent}_\mu f,$$

where ε may be taken equal to 0 with a finite c_0 if B is diagonalisable and more generally if none of the eigenvalues in $\{\Re(\lambda) = \rho\}$ is defective. We will prove (Corollary 12 below) that in fact, when N is the maximal size of a Jordan block associated to an eigenvalue in $\{\Re(\lambda) = \rho\}$ in the decomposition of B , then there exists a constant c such that

$$\operatorname{Ent}_\mu(P_t f) \leq c(1 + t^{2(N-1)})e^{-2\rho t} \operatorname{Ent}_\mu f.$$

The power $t^{2(N-1)}$ is optimal (see Proposition 13).

Apart from these revisited versions of known results, our main interest lies in the study of systems of attracting particles on a graph. This is done in Section 4. Let W be an even,

smooth, strongly convex potential on \mathbb{R}^d and let $\mathbf{X} = (X_i)_{i=0..N}$ be a chain of $N + 1$ particles interacting with their neighbours via the dynamics

$$\begin{cases} dX_0 &= -\nabla W(X_0 - X_1)dt + \sqrt{T_0}dB_t \\ dX_i &= -(\nabla W(X_i - X_{i-1}) + \nabla W(X_i - X_{i+1}))dt \quad \text{for } i = 1..(N-1) \\ dX_N &= -\nabla W(X_N - X_{N-1})dt + \sqrt{T_N}d\tilde{B}_t \end{cases}$$

where B and \tilde{B} are independant standard Brownian motion on \mathbb{R}^d and $T_0, T_N > 0$.

The process \mathbf{X} is a toy model related to the chains of hamiltonian oscillators introduced by Eckmann, Pillet and Rey-Bellet in [17] (see also the more recent [19] and references within). Except in some particular cases, the invariant measures of such processes are not explicit, and therefore it is not clear, for instance, whether they satisfy some functional inequality such as the Poincaré or log-Sobolev ones (see the discussion in [30, § 9.2 p.67]; also note that the results of [3, 13] do not apply since the dynamics is not reversible).

If $\bar{X} = \frac{1}{N+1} \sum_{i=0}^N X_i$ is the center of mass of the system, then $\mathbf{X} - \bar{X}$ is ergodic. The process \mathbf{X} is highly degenerated, since the randomness of a system in \mathbb{R}^{dN} is only given by a brownian motion on \mathbb{R}^{2d} . Nevertheless we will prove for $\mathbf{X} - \bar{X}$ (and for similar models) the invariant measure μ_N satisfies a log-Sobolev inequality

$$\text{Ent}_{\mu_N} f \leq cN^2 \int \frac{|\nabla f|^2}{f} d\mu_N \quad \forall f \in \mathcal{A} \text{ s.t. } \sum_{i=0}^N \nabla_i f = 0,$$

where c does not depend on N (the condition $\sum \nabla_i f = 0$ compensate the fact the center of mass is fixed). The N^2 is optimal since, when $W(x) = \frac{1}{2}|x|^2$, it can be checked the optimal constant in the log-Sobolev inequality is indeed of order N^2 .

General notations and conventions

The following holds throughout the whole paper. A vector $x \in \mathbb{R}^d$ will always be equated to a 1-column matrix, and therefore the scalar product in \mathbb{R}^d will indifferently be written $x.y$ or $x^T y$ where A^T denotes the transpose of a matrix A . Unless otherwise specified, $|x|$ is the euclidian norm of x , $|A|$ is the operator norm $|A| = \sup\{|Ax|, |x| = 1\}$ and $\sigma(A)$ is the spectrum of A . When Q is a quadratic operator, we still call $f, g \mapsto Q(f, g) = \frac{1}{4}(Q(f+g) - Q(f-g))$ the associated bilinear symetric operator obtained by polarization. When A and B are symetric matrices and $\rho \in \mathbb{R}$ then $A \geq B$ (resp. $A \geq \rho$) means $x^T A x \geq x^T B x$ (resp. $\geq \rho|x|^2$) for all $x \in \mathbb{R}^d$. We note $\mathcal{P}(E)$ the set of probability laws on a space E and for $\mu \in \mathcal{P}(E)$ in some cases we use the operator notation $\mu f = \int f d\mu$. When $x \mapsto b(x)$ is a smooth vector field on \mathbb{R}^d we note $J_b(x) = (\partial_i b_j(x))_{1 \leq i, j \leq d}$ the Jacobian matrix of b (the i is for the line, the j for the column).

2 Gamma Calculus

We recall here some ideas from [27]. For a bounded measurable $f : E \rightarrow \mathbb{R}$, let $P_t f(x) = \mathbb{E}(f(X_t) | X_0 = x)$ be a Markov semi-group associated to a Markov process X on a Polish space E , $Lf(x) = \lim_{t \rightarrow 0} \frac{1}{t}(P_t f(x) - f(x))$ be its generator for f in $\mathcal{D}(L)$ the domain of L . Assume $\mathcal{D}(L)$ contains a core \mathcal{A} which is an algebra fixed by L and P_t . Let $\mathcal{A}_+ = \{f \in \mathcal{A}, f \geq 0\}$ and let $\Phi : \mathcal{A}_+ \rightarrow \mathcal{A}$ be differentiable with respect to pointwise convergence topology with differential operator $d\Phi$, in the sense $\lim_{s \rightarrow 0} \frac{1}{s}(\Phi(f+sg) - \Phi(f))(x) = (d\Phi(f).g)(x)$ for all $x \in E, f, g \in \mathcal{A}_+$. For $f \in \mathcal{A}_+$ we define

$$\Gamma_{L, \Phi}(f) = \frac{1}{2}(L\Phi(f) - d\Phi(f).Lf)$$

and we suppose $\Gamma_{L,\Phi}(f) \in \mathcal{A}$. When there is no ambiguity on the generator we only write Γ_Φ , and when $\Phi(f) = f^2 + af + b$ with $(a, b) \in \mathbb{R}^2$ we retrieve the usual carré du champ operator which is simply denoted by Γ . When $\Phi(f) = \Gamma(f)$, we retrieve the classical Bakry-Émery Γ_2 operator, since for a quadratic form $Q(f)$, $dQ(f).g = 2Q(f, g)$.

Remark: The factor $\frac{1}{2}$ in the definition of $\Gamma_{L,\Phi}$ is disputable, since with this definition we will most of the time deal with $2\Gamma_{L,\Phi}$, but we wanted to be consistent with the classical definitions of Γ and Γ_2 . The $\frac{1}{2}$ in the carré du champ ensures that $\Gamma f = |\nabla f|^2$ when $L = \Delta$ is the Laplace-Beltrami of a Riemannian manifold. Having said that, the generator of a standard Brownian motion is $\frac{1}{2}\Delta$, which advocates for a definition with no $\frac{1}{2}$.

2.1 Motivations

Without claiming to be comprehensive, we want to emphasize in this section some reasons why Γ_Φ operators may be usefull and especially why it may be interesting to work with more general functions than only Γ (or functions based on Γ , such as in [9] and references within). The examples below come from other works to which we will refer for complete proofs and computations.

2.1.1 Sub-commutation Φ /semigroup

The Γ_Φ operator naturally appears in the interpolation from $\Phi(P_t f)$ to $P_t \Phi(f)$. Indeed, for $f \in \mathcal{A}_+$, $t > 0$, $x \in E$ and $s \in [0, t]$, let $\psi(s) = P_s \Phi(P_{t-s} f)(x)$. Then

$$\psi'(s) = 2P_s \Gamma_\Phi(P_{t-s} f)(x).$$

As a direct consequence, we have the following:

Lemma 1. *The so-called curvature condition $\Gamma_\Phi \geq \rho \Phi$ for some $\rho \in \mathbb{R}$, is equivalent to*

$$\forall f \in \mathcal{A}_+ \quad \Phi(P_t f) \leq e^{-2\rho t} P_t \Phi(f) \quad (1)$$

Proof. If the curvature condition holds, the Gronwall's Lemma yields $\psi(0) \leq e^{-2\rho t} \psi(t)$, which is (1). Differentiating (1) at $t = 0$ yields the curvature condition. \square

Example 1: Consider the TCP process on \mathbb{R}_+ with generator

$$Lf(x) = f'(x) + \lambda \left(\int_{h=0}^{\infty} f(hx) d\nu(h) - f(x) \right)$$

with $\lambda > 0$, an $\nu \in \mathcal{P}(\mathbb{R}_+)$. For instance with $\nu = \delta_{h_0}$, we compute $\Gamma f(x) = \frac{\lambda}{2} (f(h_0 x) - f(x))^2$ and (possibly with the help of Lemma 5 below)

$$\Gamma_2 f(x) = \frac{\lambda^2}{4} \left((f(h_0^2 x) - 2f(h_0 x) + f(x))^2 + 2(h_0 - 1)f'(h_0 x) (f(h_0 x) - f(x)) \right),$$

so that no curvature condition $\Gamma_2 \geq \rho \Gamma$ can hold. In contrast, with $\Phi(f) = |\nabla f|^2$, it is proved in [26, Proposition 15] that $\Gamma_\Phi \geq \frac{\lambda}{2} (1 - \int h^2 d\nu(h)) \Phi$ and thus

$$|\nabla P_t f|^2 \leq e^{-\lambda(1 - \int h^2 d\nu(h))t} P_t |\nabla f|^2.$$

2.1.2 Functional inequalities

In the general case, if μ is an invariant law of P_t , integrating (1) reads

$$\int \Phi(P_t f) d\mu \leq e^{-2\rho t} \int \Phi(f) d\mu.$$

We say P_t is ergodic if it admits a unique invariant law μ such that for all $f \in \mathcal{A}, x \in E$, $P_t f(x) \rightarrow \int f d\mu$ as $t \rightarrow \infty$. In this case we suppose $\mathcal{A} \subset L^2(\mu)$.

Lemma 2. *Suppose P_t is ergodic and $\Gamma_{\Phi_1}(P_t f) \leq \gamma(t) P_t \Phi_2(f)$ for all $f \in \mathcal{A}_+$, with $\gamma \in L^1(\mathbb{R}_+)$. Then*

$$\forall f \in \mathcal{A}_+, \quad \int \Phi_1(f) d\mu - \Phi_1\left(\int f d\mu\right) \leq 2 \left(\int_0^\infty \gamma(s) ds\right) \int \Phi_2(f) d\mu. \quad (2)$$

In particular if the "generalized Γ_2 condition" $\Gamma_{\Gamma_{\Phi_1}} \geq \rho \Gamma_{\Phi_1}$ holds with $\rho > 0$, then

$$\forall f \in \mathcal{A}_+, \quad \int \Phi_1(f) d\mu - \Phi_1\left(\int f d\mu\right) \leq \frac{1}{\rho} \int \Gamma_{\Phi_1}(f) d\mu.$$

Proof. If $\psi_1(s) = P_s \Phi_1(P_{t-s} f)$,

$$\begin{aligned} P_t \Phi_1(f) - \Phi_1(P_t f) &= \psi_1(t) - \psi_1(0) \\ &= 2 \int_0^t P_s \Gamma_{\Phi_1}(P_{t-s} f) ds \\ &\leq 2 \left(\int_0^t \gamma(s) ds\right) P_t \Phi_2(f). \end{aligned} \quad (3)$$

Conclusion follows when $t \rightarrow \infty$. When the generalized Γ_2 condition holds, according to Lemma 1 we may take $\Phi_2 = \Gamma_{\Phi_1}$ and $\gamma(t) = e^{-2\rho t}$. \square

Remark: An inequality of the form (3) is called a local functional inequality: it is a functional inequality satisfied by all the measures $(P_t(x))_{x \in E} = (\delta_x P_t)_{x \in E}$, uniformly in $x \in E$. By contrast an inequality of the form (2) may be called an integrated functional inequality. A local inequality may hold even if $\gamma \notin L^1(\mathbb{R}_+)$: for instance for the Brownian motion in \mathbb{R}^d , $L = \frac{1}{2}\Delta$ satisfies the curvature condition $\Gamma_2 \geq 0$, and thus $P_t f^2 - (P_t f)^2 \leq t P_t |\nabla f|^2$.

Example 2: Consider a general Ornstein-Uhlenbeck process on \mathbb{R}^d , namely a diffusion on \mathbb{R}^d with generator

$$Lf(x) = -(Bx) \cdot \nabla f(x) + \operatorname{div}(D \nabla f)(x) \quad (4)$$

where $B, D \in \mathcal{M}_{d \times d}(\mathbb{R})$, D is symmetric positive semidefinite. When $\ker D$ does not contain any non-trivial subspace which is invariant by B , the process is hypoelliptic (cf. [21, p. 148]). The carré du champ operator is $\Gamma f = (\nabla f)^T D \nabla f$ and if D is not definite positive the curvature condition $\Gamma_2 \geq \rho \Gamma$ cannot be fulfilled (in other words, the Γ_2 -curvature is $-\infty$): it would imply via Lemma 2 a log-Sobolev inequality

$$\operatorname{Ent}_\mu f \leq \frac{1}{2\rho} \int \frac{\Gamma f}{f} d\mu,$$

which is impossible since the right hand side vanishes for some non-constant f , for which the entropy is positive.

As Arnold and Erb proved in [2], and as we will see in a revisited version in Section 3, even if D is not definite positive, if all the eigenvalues of B lies in $\{\lambda \in \mathbb{C}, \Re(\lambda) > \rho\}$ for some $\rho > 0$ then the process is ergodic and there exists an explicit constant c such that

$$\frac{|\nabla P_t f|^2}{P_t f} \leq c e^{-\rho t} P_t \left(\frac{|\nabla f|^2}{f} \right).$$

For $\Phi_1(f) = f \ln f$ it is classical that $\Gamma_{\Phi_1} = \frac{(\nabla f)^T D \nabla f}{2f}$ and thus Lemma 2 yields the log-Sobolev inequality

$$\text{Ent}_\mu f \leq \frac{c|D|}{2\rho} \int \frac{|\nabla f|^2}{f} d\mu.$$

Of course, in this example the invariant law μ is gaussian and may be explicited ([2, Theorem 3.1]), and all gaussian laws satisfy such an inequality. But the arguments extend to the case

$$Lf(x) = -b(x) \cdot \nabla f(x) + \text{div}(D \nabla f)(x) \quad (5)$$

if we simply assume the Jacobian matrix of b is such that $J_b \geq B$ for some $B \in \mathcal{M}_{d \times d}(\mathbb{R})$. See Section 4 for an application.

For another example of the link between functional inequalities and Γ calculus, see the proof of the Hölder Inequality in the Preface of [5]. Or, actually, see the whole book.

2.1.3 Hypocoercivity

Some functional inequality such as those proved in Lemma 2 may also be known a priori, which leads to a third way to use Γ calculus:

Lemma 3. *Suppose P_t is ergodic and its invariant measure satisfies*

$$\forall f \in \mathcal{A}_+ \quad 0 \leq \int \Phi_1(f) d\mu - \Phi_1 \left(\int f d\mu \right) \leq c \int \Phi_2(f) d\mu.$$

If moreover $\Gamma_{\Phi_2} \geq \rho \Phi_2 - \beta \Gamma_{\Phi_1}$ for some $\beta > 0$ then, writing

$$W(t) = \beta \left(\int \Phi_1(P_t) d\mu - \Phi_1 \left(\int f d\mu \right) \right) + \int \Phi_2(P_t f) d\mu,$$

we have $W(t) \leq e^{-\frac{2\rho t}{1+\beta c}} W(0)$ and in particular

$$\int \Phi_2(P_t f) d\mu \leq (1 + \beta c) e^{-\frac{2\rho t}{1+\beta c}} \int \Phi_2(f) d\mu.$$

Proof. Since μ is invariant, $\int L\Phi_i(f) d\mu = 0$ for all f and $i = 1, 2$, and

$$\begin{aligned} W'(t) &= -2 \int (\Gamma_{\Phi_2} + \beta \Gamma_{\Phi_1})(P_t f) d\mu \\ &\leq -2\rho \int \Phi_2(P_t f) d\mu \\ &\leq -\frac{2\rho}{1 + \beta c} W(t). \end{aligned}$$

By the Gronwall's Lemma,

$$\begin{aligned} \int \Phi_2(P_t f) d\mu &\leq W(t) \\ &\leq W(0) e^{-\frac{\rho t}{1+\beta c}} \\ &\leq (1 + \beta c) e^{-\frac{\rho t}{1+\beta c}} \int \Phi_2(f) d\mu. \end{aligned}$$

□

Example 3: Consider the kinetic Fokker-Planck diffusion on \mathbb{R}^{2d} with generator

$$Lf(x, y) = \left[y \cdot \nabla_x - \left(y + \frac{1}{\varepsilon} \nabla_x U \right) \cdot \nabla_y + \Delta_y \right] f(x, y), \quad (6)$$

where $\varepsilon > 0$ and $U \in \mathcal{C}^\infty(\mathbb{R}^d)$ is quadratic at infinity (cf. [27] for precise assumptions on U and proofs of the upcoming assertions ; see also Lemma 8 below). The carré du champs is $\Gamma(f) = |\nabla_y f|^2$ and again the Γ_2 -curvature is $-\infty$.

The invariant law is $\mu = e^{-\frac{U(x)}{\varepsilon} - \frac{|y|^2}{2}} dx dy$, the Gibbs law associated to the Hamiltonian $H(x, y) = \frac{U(x)}{\varepsilon} + \frac{|y|^2}{2}$. It satisfies a log-Sobolev inequality

$$\text{Ent}_\mu(f) \leq c_\varepsilon \int \frac{|\nabla f|^2}{f} d\mu,$$

with c_ε such that $\lim_{\varepsilon \rightarrow 0} \varepsilon \ln c_\varepsilon = d_* > 0$ is the critical depth of the potential U . Let

$$M = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad \Phi_2(f) = \frac{(\nabla f)^T M \nabla f}{f} = \frac{|(\nabla_x - \nabla_y)f|^2 + |\nabla_y f|^2}{f}.$$

Under the assumption that the Hessian $\nabla_x^2 U$ is bounded, there exist $\rho, \beta > 0$ such that

$$\Gamma_{\Phi_2}(f) \geq \rho \Phi_2(f) - \beta \frac{|\nabla_y f|^2}{f}.$$

Applying Lemma 3 with $\Phi_1(f) = \text{Ent}_\mu(f)$, we obtain

$$\text{Ent}_\mu(P_t f) \leq e^{-\lambda_\varepsilon t} \left(\frac{3}{\beta} \int \frac{|\nabla f|^2}{f} d\mu + \text{Ent}_\mu(f) d\mu \right)$$

with $\lambda_\varepsilon = \frac{2\rho}{1+\beta c_\varepsilon}$. Note that β and ρ depends on ε , but not too much, in the sense we still have $\lim_{\varepsilon \rightarrow 0} \varepsilon \ln \lambda_\varepsilon = -d_*$ (see [27] for the consequences on the simulated annealing algorithm based on the kinetic Fokker-Planck dynamics).

The idea to work with twisted gradients of the form $(\nabla f)^T M \nabla f$ is due to Villani [30] and was initially applied to the kinetic Fokker-Planck process. This case had already been written in "Γ settings" (meaning, considering pointwise inequalities and quantities of the form $P_s \Phi(P_{t-s} f)$ rather than $\int \Phi(P_t f) d\mu$) by Baudoin in [7]. Later, Dobeault, Mouhot and Schmeiser proposed in [15] a similar strategy in L^2 with no space derivatives involved, which have from then proved efficient in many contexts.

These methods yield convergences of the form $V(t) \leq c e^{-\rho t} V(0)$ with $c > 1$, which are called hypocoercive: the main differences with the case $c = 1$ is that hypocoercivity is stable by equivalence (in the sense if $c_1 V(t) \leq W(t) \leq c_2 V(t)$ with $c_1, c_2 > 0$, then W is also

hypocoercive), and is not equivalent to the functional inequality " $V'(0) \leq -\rho V(0)$ for all $f \in \mathcal{A}$ " (such as the Poincaré inequality is equivalent to the exponential decay of the variance, and the log-Sobolev inequality is equivalent to the exponential decay of the entropy; see [9]).

For another application of Lemma 3, we refer to the study in [26] of the TCP with generator

$$Lf(x) = f'(x) + \lambda(x)(f(hx) - f(x)),$$

where $h \in [0, 1)$ and λ is a positive increasing function. Note this process is not a diffusion, and the invariant law μ is not explicit. Nevertheless it is possible to prove μ satisfies a log-Sobolev inequality, and the \mathcal{H}^1 norm $\int (P_t f - \int f d\mu)^2 + |\nabla P_t f|^2 d\mu$ decays exponentially fast with an explicit rate.

2.1.4 Global hypoellipticity

Lemmas 1, 2 and 3 are mostly related to the long-time behaviour of the semi-group, but Γ calculus may also reveal short-time regularization properties:

Lemma 4. *For some $t_0 \in (0, \infty]$, let $\Phi_t(f)$ be defined for $t \in [0, t_0]$, $f \in \mathcal{A}_+$ and suppose that $\partial_t \Phi_t(f)$ exists and is in \mathcal{A} for all $f \in \mathcal{A}_+$.*

a) *It $2\Gamma_{\Phi_t} f \geq \partial_t \Phi_t f$ for all $t \in [0, t_0]$, $f \in \mathcal{A}_+$, then*

$$\Phi_t(P_t f) \leq \Phi_t(\Phi_0 f). \quad (7)$$

b) *If P_t admits an invariant law μ such that $\mathcal{A} \subset L^2(\mu)$ and $2 \int \Gamma_{\Phi_t} f d\mu \geq \int \partial_t \Phi_t f d\mu$ for all $t \in [0, t_0]$, $f \in \mathcal{A}_+$, then*

$$\int \Phi_t(P_t f) d\mu \leq \int \Phi_0 f d\mu. \quad (8)$$

Proof. Let $\psi(s) = P_s \Phi_{t-s}(P_{t-s} f)$, so that $\psi'(s) = P_s (2\Gamma_{\Phi_{t-s}} - \partial_t \Phi_{t-s})(P_{t-s} f)$. In the first case $\psi'(s) \geq 0$, and in the second one $\int \psi'(s) d\mu \geq 0$, which conclude in both cases. \square

Example 4: this idea was introduced by Hérau in [20, Section 3] (see also [30, A.21 p.145]) to study the kinetic Fokker-Planck with generator (6) via

$$\Phi_t(f) = f^2 + t(a|\nabla_y f|^2 + bt|\nabla_y f + ct\nabla_x f|^2)$$

where $a, b, c \in \mathbb{R}$. Under the assumption that the Hessian $\nabla^2 U$ is bounded (see also [30, Theorem A.8, A.15 and Remark p.152] for a proof under the weaker assumption $|\nabla^2 U| \leq C(1 + |\nabla U|)$) a carefull choice of a, b and c leads to $\partial_t \int \Phi_t(P_t f) d\mu \leq 0$. In particular since $\int (P_t f)^2 d\mu \leq \int f^2 d\mu$, denoting by $\|\cdot\|$ the $L^2(\mu)$ norm,

$$\|\nabla P_t f\|^2 \leq C \frac{\|f - \mu f\|^2}{t^3}$$

for $t \leq t_0$ for some $C, t_0 > 0$.

Lemma 4 is reminiscent of the so-called reversed local entropy inequalities (see [9, Theorem 2]). In fact, suppose Φ satisfies the generalized Γ_2 condition $\Gamma_{\Gamma_\Phi} f \geq \rho \Gamma_\Phi$, $\rho \in \mathbb{R}$. Let $c(t) = \frac{e^{2\rho t} - 1}{\rho}$ if $\rho \neq 0$ and $c(t) = 2t$ if $\rho = 0$ be the solution of $c'(t) = 2(1 + \rho c(t))$ with $c(0) = 0$, and

$$\Phi_t = \Phi + c(t)\Gamma_\Phi.$$

Then $2\Gamma_{\Phi_t} \geq \partial_t \Phi_t$, hence for all $t \geq 0$

$$P_t \Phi(f) - \Phi(P_t f) \geq c(t)\Gamma_\Phi(P_t f). \quad (9)$$

2.2 Quadratic Φ 's, gradient bounds and couplings

Consider the case of $\Phi(f) = C_1 f.C_2 f$ when $C_i = (c_{i,1}, \dots, c_{i,r})$ is a linear operator from \mathcal{A} to \mathcal{A}^r for some $r \geq 1$. These Φ 's play a particular role as they behave well with any Markov generator and not only with diffusion ones. The test functions f do not need to be positive.

By convention we write $\Gamma(C_1 f, C_2 f) = \sum_{i=1}^r \Gamma(c_{1,i} f, c_{2,i} f)$ and $[L, C_i] = ([L, c_{i,1}], \dots, [L, c_{i,r}])$, where $[L, c_{i,j}] = L c_{i,j} - c_{i,j} L$ is the commutator of L and $c_{i,j}$.

Lemma 5. *If $\Phi(f) = C_1 f.C_2 f$ then*

$$\Gamma_\Phi(f) = \Gamma(C_1 f, C_2 f) + \frac{1}{2} C_1 f.[L, C_2] f + \frac{1}{2} [L, C_1] f.C_2 f$$

Proof. This is a straightforward consequence of the definitions, since $d\Phi(f)g = C_1 f.C_2 g + C_1 g.C_2 f$. \square

In particular, if $C_1 = C_2 = C$, since Γ is a positive quadratic operator, we always have

$$\Gamma_{|C|^2} f \geq C f.[L, C] f. \quad (10)$$

Example 5: Let $L f(x) = b(x).\nabla f(x)$ be a deterministic transport operator on \mathbb{R}^d . Then $\Gamma = 0$ and

$$\begin{aligned} \Gamma_{|\nabla|^2} f &= (\nabla f)^T [b.\nabla, \nabla] f \\ &= -(\nabla f)^T J_b \nabla f. \end{aligned}$$

If $J_b(x) \leq -\rho$ for all $x \in \mathbb{R}^d$ with $\rho \in \mathbb{R}$, from Lemma 1,

$$|\nabla P_t f|^2 \leq e^{-2\rho t} P_t |\nabla f|^2.$$

On the other hand if $J_b(x) \geq \rho$ for all $x \in \mathbb{R}^d$ we may invert the signs in the proof of Lemma 1 to get

$$|\nabla P_t f|^2 \geq e^{2\rho t} P_t |\nabla f|^2.$$

Here of course $P_t f(x) = f(\varphi_x(t))$ where φ is the flow solution of

$$\varphi_x(0) = x \quad \varphi'_x(t) = b(\varphi_x(t)).$$

The Jacobian matrix measures how two trajectories of the deterministic flow tends to get closer or to drift apart. Indeed,

$$\begin{aligned} \varphi_x(t) - \varphi_y(t) &= x - y + t J_b(x)(x - y) + t \underset{y \rightarrow x}{o} (x - y) + \underset{t \rightarrow 0}{o} (t) \\ \Rightarrow |\varphi_x(t) - \varphi_y(t)|^2 &= |x - y|^2 + 2t(x - y)^* J_b(x)(x - y) + t \underset{y \rightarrow x}{o} (|x - y|^2) + |x - y| \underset{t \rightarrow 0}{o} (t). \end{aligned}$$

A negative Jacobian means the flow locally contracts the space in the vicinity of all points.

In the classical Bakry-Émery theory, $\Phi f = \Gamma f$ is a natural choice in particular because it only depends on the Markov dynamics (it is, in a sense, intrinsic). On the other hand, the choice $\Phi f = |\nabla_d f|^2$, where ∇_d is the (upper-)gradient associated to the metric d on E by

$$|\nabla_d f|(x) := \lim_{r \rightarrow 0} \sup_{0 < d'(x, y) \leq r} \frac{|f(x) - f(y)|}{d(x, y)}, \quad (11)$$

is linked to the Wasserstein distances on $\mathcal{P}(E)$, as shown by Kuwada [22, 23] (see also [12]). Recall the Wasserstein distance \mathcal{W}_p on $\mathcal{P}(E)$ is defined by

$$\mathcal{W}_p(\mu, \nu) = \inf\{(\mathbb{E}(d^p(U, V)))^{\frac{1}{p}}, \text{ law}(U) = \mu, \text{ law}(V) = \nu\}.$$

When the marginal laws of (U, V) are μ and ν , (U, V) is called a coupling of μ and ν . For $\mu \in \mathcal{P}(E)$ we write μQ the image of μ by a Markov operator Q , so that $(\mu Q)f = \mu(Qf)$.

We state the following result only in \mathbb{R}^d endowed with the euclidian metric and refer to [22, 23] for more general settings:

Theorem 6 (Theorem 2.2 of [22]). *Let P be a Markov semi-group on \mathbb{R}^d and let $p > 1$ and $q = \frac{p}{p-1}$ be its Hölder conjugate. Then*

$$\forall \mu, \nu \in \mathcal{P}(\mathbb{R}^d), \quad \mathcal{W}_p(\mu P, \nu P) \leq \gamma \mathcal{W}_p(\mu, \nu) \quad (12)$$

for some $\gamma > 0$ implies

$$\forall f \in \mathcal{C}_{b,L}, x \in \mathbb{R}^d, \quad |\nabla P f|(x) \leq \gamma (P(|\nabla f|^q))^{\frac{1}{q}}(x) \quad (13)$$

where $\mathcal{C}_{b,L}$ is the set of bounded continuous Lipschitz functions on \mathbb{R}^d .

Moreover if for all $x \in \mathbb{R}^d$, $P(x)$ admits a continuous density with respect to the Lebesgue measure, then the converse is true: (13) implies (12).

Example 6: Consider the jump operator $Lf(x) = \lambda(Qf(x) - f(x))$ on \mathbb{R}^d where $Qf(x)$ is a Markov operator and $\lambda > 0$ is constant. If $(Y_k)_{k \geq 0}$ is a Markov chain with operator Q , in the sense $\mathbb{E}(f(Y_{k+1})|Y_k = y) = Qf(y)$, and $(N_t)_{t \geq 0}$ is a Poisson process with intensity λ , then $X_t = Y_{N_t}$ is a Markov process with generator L . It is possible to define two processes X and X' with generator L starting at different points x and x' that will always jump together, by considering the same Poisson process, so that the way both processes may get closer or drift apart is given by the behaviour of the associated Markov chains Y and Y' . Suppose Q is such that one can define a coupling (Y, Y') where both Y and Y' are Markov chains associated to Q and

$$\mathbb{E}(|Y_{k+1} - Y'_{k+1}|^2 | Y_k = y, Y'_k = y') \leq \gamma |y - y'|^2.$$

for some $\gamma > 0$. If $\gamma < 1$ the space is contracted by the jumps, if $\gamma > 1$ it may be expanded. In any case

$$\begin{aligned} \mathbb{E}(|X_t - X'_t|^2) &\leq \mathbb{E}(\gamma^{N_t}) |x - x'|^2 \\ &= e^{\lambda(\gamma-1)t} |x - x'|^2. \end{aligned}$$

The above couplings prove that for all $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\begin{aligned} \mathcal{W}_2^2(\mu Q, \nu Q) &\leq \gamma \mathcal{W}_2^2(\mu, \nu) \\ \mathcal{W}_2^2(\mu P_t, \nu P_t) &\leq e^{-\lambda(\gamma-1)t} \mathcal{W}_2^2(\mu, \nu) \end{aligned}$$

which implies by Theorem 6 that for all $f \in \mathcal{C}_{b,L}$

$$|\nabla Q f|^2 \leq \gamma Q |\nabla f|^2 \quad (14)$$

$$|\nabla P_t f|^2 \leq e^{-\lambda(\gamma-1)t} P_t |\nabla f|^2. \quad (15)$$

With Γ calculus, we directly obtain (15) from (14):

$$\begin{aligned}\Gamma_{L,|\nabla \cdot|^2} f &= \frac{\lambda}{2} \left(Q|\nabla f|^2 + |\nabla f|^2 - 2(\nabla f)^T (\nabla Q f) \right) \\ &\geq \frac{\lambda}{2} \left(Q|\nabla f|^2 + |\nabla f|^2 - 2|\nabla f| \sqrt{\gamma Q |\nabla f|^2} \right) \quad \text{given (14)} \\ &\geq \frac{\lambda(1-\gamma)}{2} |\nabla f|^2\end{aligned}$$

and Lemma 1 yields (15). Note that $P_t(x)$ is not absolutely continuous with respect to the Lebesgue measure, since $\mathbb{P}(X_t = x \mid X_0 = x) = e^{-\lambda t}$, and thus we cannot retrieve the Wasserstein contraction from Theorem 6.

For an example where the metric is not the euclidian one, we refer to the study of the TCP with linear rate on \mathbb{R}_+ , with generator

$$Lf(x) = f'(x) + x(f(hx) - f(x))$$

where $h \in [0, 1)$. A coupling approach may be found in [6], and a Γ one in [26, Section 4.4].

2.3 Diffusions and entropic Φ 's

In this subsection we suppose \mathcal{A} is fixed by the composition by any smooth compactly supported $a \in \mathcal{C}_c^\infty(\mathbb{R})$, and by exponentiation $f \rightarrow f^\alpha$ for all $\alpha \geq 0$.

We say L is a diffusion operator (or equivalently $(P_t)_{t \geq 0}$ is a diffusion semi-group or $(X_t)_{t \geq 0}$ a diffusion process) if one of the following equivalent conditions holds:

(i) for all $f, g \in \mathcal{A}$,

$$L(fg) = gLf + fLg + 2\Gamma(f, g)$$

(ii) for all $a \in \mathcal{C}_c^\infty$, $f \in \mathcal{A}$,

$$L(a(f))(x) = a'(f(x))Lf(x) + a''(f(x))\Gamma f(x).$$

(iii) for all $a \in \mathcal{C}_c^\infty$, $f, g \in \mathcal{A}$

$$\Gamma(a(f), g) = a'(f)\Gamma(f, g)$$

When L is a diffusion then conditions (i) and (ii) hold for any a such that the tems are well-defined for $f \in \mathcal{A}$. As a direct consequence of the condition (ii), if $\Phi f(x) = a(f(x))$ and L is a diffusion operator, $\Gamma_\Phi f = \frac{a''(f)}{2}\Gamma(f)$. More generally,

Lemma 7. *If L is a diffusion operator and $\Phi_2(f)(x) = a(\Phi_1 f(x))$ then*

$$\Gamma_{\Phi_2} f = a'(\Phi_1 f)\Gamma_{\Phi_1} f + \frac{a''(\Phi_1 f)}{2}\Gamma(\Phi_1 f)$$

Proof. It results from the definitions and the fact $d(a(\Phi_1)) = a'(\Phi_1)d\Phi_1$. \square

Example 7: If $\Phi_1 \geq 0$ satisfies the curvature condition $\Gamma_{\Phi_1} \geq \rho\Phi_1$ then for all $p > 1$, from Lemma 7, $\Phi_2 = \Phi_1^p$ satisfies $\Gamma_{\Phi_2} \geq p\rho\Phi_2$, which yields $\Phi_1(P_t f) \leq e^{-\rho t} (P_t(\Phi_1 f)^p)^{\frac{1}{p}}$. Of course by Jensen inequality this is weaker than $\Phi_1(P_t f) \leq e^{-\rho t} P_t \Phi_1 f$. When $\Phi_1 f = |\nabla f|^2$, considering Theorem 6, this is consistent with the fact a contraction of \mathcal{W}_p is stronger than a contraction of $\mathcal{W}_{p'}$ when $p' < p$.

Definition. When a is a convex function on \mathbb{R}_+ and $\nu \in \mathcal{P}(E)$,

$$\text{Ent}_\nu^a(f) = \nu(a(f)) - a(\nu f),$$

which is positive, is called the a -entropy of f with respect to ν . For $E = \mathbb{R}^d$ we will say $a \in \mathcal{C}^4$ is an admissible function if $\frac{1}{a''}$ is positive concave.

Note that an admissible function is necessarily strictly convex, and thus $\text{Ent}_\nu^a(f) = 0$ iff ν -almost everywhere $f = \nu f$. When a is an admissible function, the curvature condition $\Gamma_2 \geq \rho\Gamma$ implies $\Gamma_{\Gamma_\Phi} \geq \rho\Gamma_\Phi$ with $\Phi f = a(f)$ (we refer to [9, Theorem 2 and Remark 3] and references within for more considerations on this matter) which, following the proof of Lemma 2, yields

$$\text{Ent}_{P_t(x)}^a(f) \leq \frac{1 - e^{-2\rho t}}{\rho} P_t(a''(f)\Gamma(f))(x).$$

When the curvature condition does not hold with Γ_2 , we can try to compute Γ_Φ for Φ of a form similar to $a''(f)\Gamma f$, which leads to the following:

Lemma 8. If L is a diffusion operator, $C : \mathcal{A} \rightarrow \mathcal{A}^r$ is a linear operator and $a : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an admissible function such that $\Phi f = a''(f)|Cf|^2 \in \mathcal{A}_+$ for all $f \in \mathcal{A}_+$, then

$$\Gamma_\Phi f \geq a''(f)Cf.[L, C]f. \quad (16)$$

Proof. Let $\Phi_1 f = |Cf|^2$ and $\alpha = \frac{1}{a''}$. By the diffusion property,

$$L\left(\frac{\Phi_1(f)}{\alpha(f)}\right) = \frac{L\Phi_1 f}{\alpha(f)} + \Phi_1 f \left(\frac{-\alpha'(f)Lf}{\alpha^2(f)}\right) + \left(\frac{1}{\alpha}\right)'' \Gamma f + 2\Gamma\left(\Phi_1 f, \frac{1}{\alpha(f)}\right)$$

with $\left(\frac{1}{\alpha}\right)'' = -\frac{\alpha''}{\alpha^2} + 2\frac{(\alpha')^2}{\alpha^3} \geq 2\frac{(\alpha')^2}{\alpha^3}$. On the other hand

$$d\left(\frac{\Phi_1}{\alpha}\right)(f).Lf = \frac{d\Phi_1(f).Lf}{\alpha(f)} - \Phi_1 f \frac{\alpha'(f)Lf}{\alpha^2(f)}.$$

Therefore

$$\Gamma_\Phi f \geq \frac{\Gamma_{\Phi_1}(f)}{\alpha(f)} + \frac{(\alpha')^2(f)}{\alpha^3(f)}\Phi_1 f + \Gamma\left(\Phi_1 f, \frac{1}{\alpha(f)}\right)$$

By Lemma 5, $\Gamma_{\Phi_1} f = \Gamma(Cf) + Cf.[L, C]f$, and by the diffusion property

$$\Gamma\left(\Phi_1 f, \frac{1}{\alpha(f)}\right) = \frac{\alpha'(f)2Cf.\Gamma(Cf, f)}{\alpha^2(f)}$$

where by convention $\Gamma(Cf, f) = (\Gamma(c_1 f), \dots, \Gamma(c_r f, f))$. As a positive bilinear form Γ satisfies the Cauchy-Schwarz inequality $|\Gamma(Cf, f)|^2 \leq \Gamma(Cf)\Gamma f$, and thus

$$\begin{aligned} \Gamma_\Phi f &\geq \frac{\Gamma(Cf) + Cf.[L, C]f}{\alpha(f)} + \frac{(\alpha')^2(f)}{\alpha^3(f)}\Phi_1 f - 2\sqrt{\frac{(\alpha')^2(f)\Phi_1(f)\Gamma f}{\alpha^3(f)}}\sqrt{\frac{\Gamma(Cf)}{\alpha(f)}} \\ &\geq \frac{Cf.[L, C]f}{\alpha(f)}. \end{aligned}$$

□

Example 8: Let b be a smooth vector field on \mathbb{R}^d , $D = Q^T Q$ be a constant semidefinite positive matrix and

$$Lf = b(x) \cdot \nabla f + \operatorname{div}(D \nabla f).$$

The carré du champ is $\Gamma f = (\nabla f)^T D \nabla f = |Q \nabla f|^2$. For $M = P^T P$ a constant positive matrix, let

$$\Phi_M(f) = a''(f) |P \nabla f|^2 = a''(f) (\nabla f)^T M \nabla f.$$

Since $[L, P \nabla] = -P J_b \nabla$, Lemma 8 yields

$$\Gamma_{\Phi_M} f \geq -a''(f) (\nabla f)^T M J_b \nabla f \quad (17)$$

This explains the computations of Examples 2 and 3. For Ornstein-Uhlenbeck processes, see also Corollary 12 below. But above all, it is the key ingredient of the next section.

3 General results

We propose now to prove the Γ counterparts of Villani's general results [30, Theorem A.15 p.158 and Theorem 28, p.42] concerning (entropic) hypoellipticity and hypocoercivity for diffusions operators.

Suppose the diffusion generator L on \mathbb{R}^d is on Hörmander form

$$L = B_0 + \sum_{i=1}^d B_i^2$$

where $B = (B_j)_{0 \leq j \leq r} = (b_j \cdot \nabla)_{0 \leq j \leq r}$ is a derivation operator on \mathbb{R}^d with \mathcal{C}^∞ coefficient b_j 's which are not necessarily linearly independant. Let a be an admissible function, that is to say a stricly convex \mathcal{C}^4 function from \mathbb{R}_+ to \mathbb{R}_+ such that $\frac{1}{a''}$ is concave. Let \mathcal{A} be the set of \mathcal{C}^∞ functions on \mathbb{R}^d whose all derivatives grow at most as polynomials at infinity. Suppose \mathcal{A} is fixed by P_t , L and that if \mathcal{A}_p is the set of functions in \mathcal{A} which are bounded by a positive constant, $f \in \mathcal{A}_p$ implies $a(f) \in \mathcal{A}$. Moreover suppose that if $\mu \in \mathcal{P}(\mathbb{R}^d)$ is invariant for P_t and have a finite exponential moment then for all positive $f \in L^1(\mu)$ such that $\operatorname{Ent}_\mu^a f < \infty$, there exists a sequence $(f_m)_{m \in \mathbb{N}}$ of Lipschitz bounded functions in \mathcal{A}_p such that as m goes to infinity, $f_m \rightarrow f$ in $L^1(\mu)$ and for all $t \geq 0$, $\operatorname{Ent}_\mu^a P_t f_m \rightarrow \operatorname{Ent}_\mu^a P_t f$ (this assumption is clearly satisfied if $a(f) = f^p$, $p > 1$, since in $L^p(\mu)$, \mathcal{A} is dense and P_t is continuous).

For $i \geq 0$, $n \in \mathbb{N}$, we write

$$\begin{aligned} C_i &= (c_{i,1} \cdot \nabla, \dots, c_{i,n} \cdot \nabla) \\ R_i &= (r_{i,1} \cdot \nabla, \dots, r_{i,n} \cdot \nabla) \end{aligned}$$

where the $r_{i,j}$'s and $c_{i,j}$'s are \mathcal{C}^∞ vector fields on \mathbb{R}^d , and $Z_i = (z_{i,k,l})_{1 \leq k,l \leq n} \in \mathcal{M}_n(\mathbb{R})$ where the $z_{i,k,l}$'s are \mathcal{C}^∞ scalar fields. By convention $R_1^T R_2$ will stand for the quadratic operator $f \mapsto (R_1 f)^T R_2 f$, and recall we set

$$[L, C_i] := ([L, c_{i,1}], \dots, [L, c_{i,n}]).$$

Theorem 9. Suppose there exist $N_c \in \mathbb{N}$ and $\lambda, \Lambda, m > 0$ such that for $i \in \llbracket 0, N_c + 1 \rrbracket$ there exist derivation operators C_i and R_i and a matrix field Z_i , all with coefficients in \mathcal{A} , satisfying:

- (i) $C_{N_c+1} = 0$, and $[B_0, C_i] = Z_{i+1} C_{i+1} + R_{i+1}$ for all $i \in \llbracket 0, N_c \rrbracket$.

(ii) $[B_j, C_i] = 0$ for all $i \in \llbracket 0, N_c \rrbracket$, $j \in \llbracket 1, d \rrbracket$,

(iii) $\lambda \leq \frac{Z_i + Z_i^T}{2} \leq \Lambda$ for all $i \in \llbracket 0, N_c \rrbracket$,

(iv) $C_0^T C_0 \leq m \sum_{j \geq 1} B_j^T B_j$ and $R_i^T R_i \leq m \sum_{j < i} C_j^T C_j$ for all $i \in \llbracket 0, N_c + 1 \rrbracket$.

Then there exists $c > 0$ such that for all $f \in \mathcal{A}_p$, $x \in \mathbb{R}^d$ and $i \in \llbracket 0, N_c \rrbracket$

$$a''(P_t f(x)) |C_i \nabla P_t f|^2(x) \leq c(1 - e^{-t})^{-2i-1} \text{Ent}_{P_t(x)}^a f. \quad (18)$$

Theorem 10. Under the same assumptions as Theorem 9, if moreover there exist $\rho, K > 0$ and a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that

$$(a) \sum_{i \geq 0} C_i^T C_i \geq \rho,$$

(b) μ is invariant for P_t and satisfies the entropic inequality

$$\forall f \in \mathcal{A}_p, \quad \text{Ent}_{\mu}^a f \leq K \int a''(f) |\nabla f|^2 d\mu, \quad (19)$$

then there exist $\kappa > 0$ such that for all $t > 0$ and for all f with $\text{Ent}_{\mu}^a f < \infty$,

$$\text{Ent}_{\mu}^a(P_t f) \leq e^{-\kappa t(1-e^{-t})^{2N_c}} \text{Ent}_{\mu}^a f. \quad (20)$$

Remark: Note that these are not exactly the same results as in [30]: the commutation condition (ii) we require is a very strong assumption (reminescent of [30, Remark 33 p. 45]). It holds if the matrix diffusion and the C_i 's are constant, or in dimension 2 if $B = (B_1, 0)$ and $C_1 = B_1$, and this is basically all. That being said for most of the models we have in mind (and to which Villani's method have been applied, to our knowledge) the diffusion matrix is indeed constant.

On the plus side our results do not need to consider adjoint operators in $L^2(\mu)$ where μ is the invariant measure (which is the case in Villani's work but also for the method of Dolbeault, Mouhot and Schmeiser), which would require to have an explicit expression or at least some informations on the density $\mu(x)$ (see discussion [30, §9.2 p.67]), or to work with the wrong invariant measure as in [10]. By contrast, Theorem 9 does not even need an invariant measure, while Theorem 10 only needs its existence and the functional inequality (19) (which may be hard to establish, of course, when μ is not explicit; see [26] for such an example, with no hypoellipticity).

Both theorems are based on the following computation:

Lemma 11. Let $\alpha(t) = 1 - e^{-t}$, $(\varepsilon_i)_{i \in \llbracket 0, N_c \rrbracket} \in (0, 1)^{N_c+1}$ and

$$\begin{aligned} \Phi_0 f &= \varepsilon_0^2 a(f) \\ \Phi_i f &= a''(f) \varepsilon_i \alpha^{2i-1}(t) | (C_{i-1} + \varepsilon_i \alpha(t) C_i) f |^2 \quad \text{for } i \in \llbracket 1, N_c \rrbracket. \end{aligned}$$

There exist $b_1, b_2, b_3 > 0$ and $\varepsilon_* \in (0, 1)$ such that for all $i \in \llbracket 0, N_c \rrbracket$, if $\varepsilon_i < \varepsilon_*$,

$$\frac{(2\Gamma_{\Phi_i} - \partial_t \Phi_i)}{a''(f)} \geq -b_1 \left(\sum_{j < i} \alpha^{2j} |C_j f|^2 \right) + b_2 \varepsilon_i^2 \alpha^{2i} |C_i f|^2 - b_3 \alpha^{2(i+1)} \varepsilon_i^4 |C_{i+1} f|^2.$$

Proof. The case $i = 0$ is clear, since $\sum_{j<0} |C_j f|^2 = 0$ and since from Lemma 7

$$2\Gamma_{\Phi_0} = a''(f)\varepsilon_0^2 \Gamma f \geq \frac{\varepsilon_0^2}{m} a''(f) |C_0 f|^2.$$

For $i \geq 1$, from Lemma 8, since $[L, C_i] = Z_{i+1}C_{i+1} + R_{i+1}$,

$$\begin{aligned} \frac{(2\Gamma_{\Phi_i} - \partial_t \Phi_i)}{a''(f)} &\geq \varepsilon_i (C_{i-1} + \alpha \varepsilon_i C_i)^T \left[2\alpha^{2i-1} (Z_i C_i + R_i) - 2\alpha^{2i-1} \alpha' \varepsilon_i C_i \right. \\ &\quad \left. + 2\alpha^{2i} \varepsilon_i (Z_{i+1} C_{i+1} + R_{i+1}) - (2i-1) \alpha' \alpha^{2i-2} (C_{i-1} + \alpha \varepsilon_i C_i) \right] \end{aligned} \quad (21)$$

In order to describe how we bound this expression, the following array $\{q_{i,j}, i = 1, 2, j = 1..5\}$ is built as follow: at line i and column j , $q_{i,j}$ is the coefficient in the left-hand side of (21) of the product $c_j^T l_i$ (the coefficient of $C_i^T C_{i-1} = C_{i-1}^T C_i$ is distributed between $c_1^T l_2$ and $c_2^T l_1$). Moreover the decomposition $q_{i,j} = 2p_{i,j} \times p'_{i,j}$ with the \times sign means that we bound $q_{i,j}(c_j f)^T (l_i f) \geq -|p_{i,j} l_i f|^2 - |p'_{i,j} c_j f|^2$ (the left term goes with l_i , the right one with c_j).

	$c_1 = C_{i-1}$	$c_2 = C_i$
$l_1 = C_{i-1}$	$-\varepsilon_i (2i-1) \alpha' \alpha^{2i-2}$	$-2((2i-1) \alpha' \alpha^{i-2}) \times (\varepsilon_i^2 \alpha^i)$
$l_2 = C_i$	$2\left(\sqrt{\frac{\lambda}{2}} \alpha^i \varepsilon_i\right) \times \left(\sqrt{\frac{2}{\lambda}} \alpha^{i-1} (Z_i - \varepsilon_i \alpha')\right)$	$\alpha^{2i} \varepsilon_i^2 (2Z_i - (2i+1) \alpha' \varepsilon_i)$
$l_3 = C_{i+1}$	$2(\alpha^{i+1} \varepsilon_i^2) \times (\alpha^{i-1} Z_{i+1})$	$2\left(\sqrt{\frac{2}{\lambda}} Z_{i+1} \varepsilon_i^2 \alpha^{i+1}\right) \times \left(\sqrt{\frac{\lambda}{2}} \varepsilon_i \alpha^i\right)$
$l_4 = R_i$	$2(\varepsilon_i \alpha^i) \times (\alpha^{i-1})$	$2(\alpha^i) \times (\alpha^i \varepsilon_i^2)$
$l_5 = R_{i+1}$	$2(\varepsilon_i^2 \alpha^i) \times (\alpha^i)$	$2\left(\alpha^{i+1} \varepsilon_i^{\frac{3}{2}}\right) \times \left(\alpha^i \varepsilon_i^{\frac{3}{2}}\right)$

For instance (line 3 column 1) in (21) appears the term $2\alpha^{2i} \varepsilon_i^2 (C_{i-1} f)^T Z_{i+1} C_{i+1} f$, which is bounded below by $-\alpha^{2(i+1)} \varepsilon_i^4 |C_{i+1} f|^2 - \alpha^{2(i-1)} \Lambda^2 |C_{i-1} f|^2$.

From the operation presented via the array, together with $\alpha, \alpha', \varepsilon_i \leq 1$, $\lambda \leq \frac{Z_i + Z_i^T}{2} \leq \Lambda$, $|R_i f|^2 \leq m \sum_{j<i} |C_j f|^2$ and $|R_{i+1} \nabla f|^2 \leq m \left(|C_i f|^2 + \sum_{j<i} |C_j f|^2 \right)$, we obtain

$$\frac{(2\Gamma_{\Phi_i} - \partial_t \Phi_i)}{a''(f)} \geq -d_1 \alpha^{2(i-1)} \sum_{j<i} |C_j f|^2 + d_2 \alpha^{2i} \varepsilon_i^2 |C_i f|^2 - d_3 \varepsilon_i^4 \alpha^{2(i+1)} |C_{i+1} f|^2$$

with some d_1, d_3 and $d_2 = \lambda \left(2 - \frac{1}{2} - \frac{1}{2}\right) + \mathcal{O}(\varepsilon_i)$ where the negligible term is uniform with respect to $i \in \llbracket 1, N_c \rrbracket$, which concludes. \square

Proof of Theorem 9. Keep the notations of Lemma 11, and let $\varepsilon_{N_c} = \frac{1}{2} \min \left(\varepsilon_*, \frac{b_2}{2b_1 + b_3} \right)$, $\varepsilon_{i-1} = \varepsilon_i^3$ for $i \leq N_c$, $\lambda_0 = 1$, and $\lambda_{i+1} = \varepsilon_i^3 \lambda_i$ for $i \geq 0$. Note in particular that

$$\sum_{j>i} \lambda_j \leq \lambda_{i+1} \sum_{k \geq 0} \frac{1}{2^k} = 2\varepsilon_i^3 \lambda_i.$$

Let

$$\Phi_{(t)}f = \sum_{i=0}^{N_c} \lambda_i \Phi_i f,$$

so that

$$\begin{aligned} \frac{2\Gamma_{\Phi_{(t)}} - \partial_t \Phi_{(t)}}{a''(f)} &\geq \sum_{i=0}^{N_c} \alpha^{2i} |C_i f|^2 \left(b_2 \lambda_i \varepsilon_i^2 - b_1 \left(\sum_{j>i} \lambda_j \right) - b_3 \lambda_{i-1} \varepsilon_{i-1}^4 \right) \\ &\geq \sum_{i=0}^{N_c} \alpha^{2i} |C_i f|^2 \lambda_i (b_2 \varepsilon_i^2 - (2b_1 + b_3) \varepsilon_i^3) \\ &\geq \frac{b_2}{2} \sum_{i=0}^{N_c} \alpha^{2i} |C_i f|^2 \lambda_i \varepsilon_i^2 > 0. \end{aligned}$$

It means $\psi(s) = P_s \Phi_{(t-s)}(P_{t-s}f)$ is increasing, and $\psi(t) \geq \psi(0)$ reads

$$\begin{aligned} P_t a(f) - a(P_t f) &\geq \frac{1}{\varepsilon_0^2 \lambda_0} \sum_{i \geq 1} \lambda_i \Phi_{(t)}(P_t f) \\ &\geq \frac{1}{c} \sum_{i=0}^{N_c} (1 - e^{-t})^{2i+1} |C_i P_t f|^2 \end{aligned}$$

for some $c > 0$. □

Proof of Theorem 10. With $\Phi_{(t)}$ defined above and the new assumptions,

$$\frac{2\Gamma_{\Phi_{(t)}} - \partial_t \Phi_{(t)}}{a''(f)} \geq \rho_1 (1 - e^{-t})^{2N_c} \left(|\nabla f|^2 + \sum_{i=0}^{N_c} |C_i f|^2 \right)$$

for some $\rho_1 > 0$. It means, writing $W(t) = \varepsilon_0^{-2} \int \Phi_{(t)}(P_t f) d\mu - a(\int f d\mu)$, that

$$\begin{aligned} W'(t) &\leq -\varepsilon_0^{-2} \rho_1 (1 - e^{-t})^{2N_c} \int \left(|\nabla P_t f|^2 + \sum_{i=0}^{N_c} |C_i P_t f|^2 \right) d\mu \\ &\leq -\rho_2 (1 - e^{-t})^{2N_c} W(t) \end{aligned}$$

for some ρ_2 thanks to Inequality (19). Therefore

$$\begin{aligned} \text{Ent}_\mu^a(P_t f) &\leq W(t) \\ &\leq e^{-\rho_2 \int_0^t (1-e^{-s})^{2N_c} ds} W_0 \\ &= e^{-\rho_2 \int_0^t (1-e^{-s})^{2N_c} ds} \text{Ent}_\mu^a f. \end{aligned}$$

A priori the result holds for $f \in \mathcal{A}_p$, but it does not depend on the regularity of f nor on its positive bound and we conclude by a density argument (indeed Inequality (19) implies μ satisfies a Poincaré inequality, which is the case $a(f) = f^2$, which in turn implies μ admits a finite exponential moment). □

Remark: a thorough reading of the proofs show that, if $\rho, \lambda \leq 1 \leq \Lambda, K, m$ (which can always be assumed), one can choose $\varepsilon_* = \frac{\lambda}{10mN_c}$, $b_1 = 7 \left(N_c^2 + \frac{\Lambda^2}{\lambda} + m \right)$, $b_2 = \frac{\lambda}{2}$ and $b_3 = \frac{3\Lambda^2}{\lambda}$

in Lemma 11, and the constants c and κ respectively in Theorems 9 and 10 may be chosen has

$$\begin{aligned} c_* &= \left(\frac{100}{\lambda} \left(N_c^2 + \frac{\Lambda^2}{\lambda} + m \right) \right)^{20N_c^2} \\ \kappa_* &= \frac{\rho}{c_* K} \end{aligned}$$

For instance, suppose we consider a family $(L_\varepsilon)_{\varepsilon>0}$ of generator such that Theorems 9 and 10 holds with N_c uniform w.r.t ε , constants $\Lambda_\varepsilon, m_\varepsilon, \lambda_\varepsilon^{-1}, \rho_\varepsilon^{-1}$ that grow at most polynomially and $\varepsilon \ln K_\varepsilon \rightarrow E > 0$ as $\varepsilon \rightarrow 0$, (which is typically the case in a metastable context). Then $\varepsilon \ln \kappa_\varepsilon \rightarrow -E$, or in other words the speed of convergence for small ε (in a large deviation scaling) is given by the degree of metastability, just as in the classical case (see [27] about this).

Application to Ornstein-Uhlenbeck processes

Corollary 12. *Let*

$$Lf(x) = -(Bx) \cdot \nabla f + \operatorname{div}(D \nabla f)$$

be the generator of an Ornstein-Uhlenbeck process. Suppose $\ker D$ does not contain any non trivial subspace which is invariant by B^T , and let M be the number of Lie brackets necessary to fulfill Hörmander's condition. Suppose $\rho = \inf\{\Re(\lambda) = \rho, \lambda \in \sigma(B)\} > 0$, and let N be the maximal dimension of a Jordan block of B corresponding to an eigenvalue in $\{\lambda \in \sigma(B), \Re(\lambda) = \rho\}$.

Then the process admits a unique invariant measure μ , and there exist constants $c, \kappa > 0$ such that for all admissible function a and all f with $\operatorname{Ent}_\mu^a f < \infty$,

$$\operatorname{Ent}_\mu^a(P_t f) \leq \min \left(c(1 + t^{2(N-1)})e^{-2\rho t}, e^{-\kappa t(1-e^{-t})^{2M}} \right) \operatorname{Ent}_\mu^a f$$

Proof. According to [2, Lemma 2.3], the condition on $\ker D$ and B is equivalent to

$$\sum_{k=0}^M B^k D (B^T)^k \geq r$$

for some $r > 0$. Let $C_0 = (u_1 \cdot \nabla, \dots, u_d \cdot \nabla) := Q \nabla$ where the u_i 's are the line of a matrix Q such that $Q^T Q = D$. Then $|C_0 f|^2 = (\nabla f)^T Q^T Q \nabla f = \Gamma f$. For $i \in \llbracket 0, M-1 \rrbracket$, simply let $Z_{i+1} = 1$, $R_{i+1} = 0$ and by induction $C_{i+1} = [L, C_i] = [-(Bx) \cdot \nabla, C_i] = Q B^{i+1} \nabla$. Let $Z_{M+1} = 1$, $C_{M+1} = 0$ and $R_{M+1} = Q B^{M+1} \nabla$, so that

$$|R_{M+1} f|^2 = (\nabla f)^T (B^{M+1})^T D B^{M+1} \nabla f \leq |Q B^{M+1}|^2 |\nabla f|^2 \leq \frac{|Q B^{M+1}|^2}{r} \sum_{j=0}^M |C_j f|^2.$$

All the conditions of Theorem 9 are fulfilled.

For $\lambda \in \sigma(B)$ let $u_m \in \ker(B^T - \lambda I)^m \setminus \ker(B^T - \lambda I)^{m-1}$ be a generalized eigenvalue of B^T associated to λ of order $m \geq 1$, and for $i \in \llbracket 1, m-1 \rrbracket$ let $u_{m-i} = (B^T - \lambda I)u_{m-i+1}$. Let $v_t = \sum_{k=1}^m \frac{t^{m-k}}{(m-k)!} u_k$, and

$$\Phi_t f = a''(f) |v_t \cdot \nabla f|^2 = a''(f) (\nabla f)^T \frac{v_t v_t^T + \bar{v}_t \bar{v}_t^T}{2} \nabla f.$$

Since $B^T v_t = \lambda v_t + v'_t$, Inequality (17) reads

$$\begin{aligned} 2\Gamma_{\Phi_t} f - \partial_t \Phi_t &\geq a''(f)(\nabla f)^T [(v_t v_t^T + \bar{v}_t \bar{v}_t^T)B - v_t(v'_t)^T + \bar{v}_t(\bar{v}'_t)^T] \nabla f \\ &= 2\Re(\lambda)\Phi_t f. \end{aligned}$$

Writing $\psi(s) = P_s \Phi_{t-s}(P_{t-s} f)$ it means $\psi'(s) \geq 2\Re(\lambda)\psi(s)$ and so

$$a''(P_t f)|v_t \cdot \nabla P_t f|^2 \leq e^{-2\Re(\lambda)t} P_t (a''(f)|u_m \cdot \nabla f|^2).$$

By induction, $|u_m \cdot \nabla f|^2 \leq |v_t \cdot \nabla f|^2 + \sum_{k=1}^{m-1} \left(\frac{t^{m-k}}{(m-k)!} \right)^2 |u_k \cdot \nabla f|^2$ yields

$$a''(P_t f)|u_m \cdot \nabla P_t f|^2 \leq c_m (1 + t^{2(m-1)}) e^{-2\Re(\lambda)t} P_t \left(a''(f) \sum_{k \leq m} |u_k \cdot \nabla f|^2 \right)$$

for some constant c_m . Finally, considering all generalized eigenspaces, it means there exists $\tilde{c} > 0$ such that

$$a''(P_t)|\nabla P_t f|^2 \leq \tilde{c}(1 + t^{2(N-1)}) e^{-2\Re(\lambda)t} P_t (a''(f)|\nabla f|^2). \quad (22)$$

A non-degenerated Gaussian invariant measure μ of the semigroup can be explicated (see [2, Lemma 3.3]). Theorem 9 implies the semigroup admits a smooth density with respect to the Lebesgue measure. Together with the contraction (22) for $a(f) = f^2$ and Theorem 6, it means for t large enough P_t is a contraction of the Wasserstein space \mathcal{W}_2 which is complete ([8]). The fixed point Theorem ensures that μ is in fact the unique invariant law and that for all initial law ν , $(\nu P_t) \xrightarrow{\mathcal{W}_2} \mu$ which mean by the Kantorovitch-Rubinstein duality that $P_t f(x) \rightarrow \mu f$ for all $x \in \mathbb{R}^d$ at least for any Lipschitz f , but then for all $f \in \mathcal{A}_p$ by a density argument since μ has a finite exponential moment. Therefore we may apply Lemma 2 with $\Phi_1 f = a(f)$ and $\Phi_2 = a''(f)|\nabla f|^2$ to obtain that

$$\forall f \in \mathcal{A}_p \quad \text{Ent}_\mu^a f \leq K \int a''(f)|\nabla f|^2 d\mu$$

holds for some $K > 0$. All the assumptions of Theorem 10 holds, and thus there exists $\kappa > 0$ such that for all $f \in \mathcal{A}_p$

$$\text{Ent}_\mu^a(P_t f) \leq e^{-\kappa t(1-e^{-t})^{2M}} \text{Ent}_\mu^a f$$

On the other hand

$$\begin{aligned} \text{Ent}_\mu^a(P_t f) &\leq K \int a''(P_t f)|\nabla P_t f|^2 d\mu \\ &\leq K e^\rho \tilde{c}(1 + t^{2(N-1)}) e^{-2\Re(\lambda)t} \int a''(P_1 f)|\nabla P_1 f|^2 d\mu \\ &\leq c(1 + t^{2(N-1)}) e^{-2\Re(\lambda)t} \text{Ent}_\mu^a f \end{aligned}$$

for some $c > 0$, where the last line is due to Theorem 9. \square

Remark: It means the short and long time behaviours of the "distance" to equilibrium

$$d_a(P_t, \mu) := \sup \{ \text{Ent}_\mu^a(P_t f), \text{Ent}_\mu^a f = 1 \}$$

are at least of order

$$1 - d_a(P_t, \mu) \underset{t \rightarrow 0}{\sim} t^{2M+1} \quad d_a(P_t, \mu) \underset{t \rightarrow \infty}{\sim} t^{2(N-1)} e^{-2\rho t}.$$

An interpretation would be the following: at small times the law of the process instantaneously approach its equilibrium by local smoothing through diffusion, and then long-range averaging is essentially due to the drift part of the dynamics.

The dependency in ρ , N and M is optimal: indeed, consider the kinetic Fokker-Planck process with generator

$$\begin{aligned} Lf(x, y) &= \left(y \partial_x - \left(\frac{1}{4}x + y \right) \partial_y + \partial_y^2 \right) f(x, y) \\ &= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{4} \\ 1 & -1 \end{pmatrix} \nabla f + \operatorname{div} \left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \nabla f \right), \end{aligned}$$

so that $\rho = \frac{1}{2}$, $N = 2$ and $M = 1$. Gadat and Miclo computed in [18, Theorem 3] the explicit value of the (squared) operator norm of $P_t - \gamma$, namely $d_a(P_t, \mu)$ with $a(f) = f^2$, which is

$$\begin{aligned} \|P_t - \mu\|^2 &= \left(1 + \frac{t^2}{2} + t \sqrt{1 + \left(\frac{t}{2} \right)^2} \right) e^{-t} \\ &\simeq 1 - \frac{t^3}{12} \quad \text{as } t \rightarrow 0 \\ &\simeq t^2 e^{-t} \quad \text{as } t \rightarrow \infty. \end{aligned}$$

In fact, as far as the long-time behaviour is concerned, this is a general fact:

Proposition 13. *In the settings of Corollary 12, let $\|P_t - \mu\| = \sup\{\|P_t f - \mu f\|_{L^2(\mu)}, \|f\|_{L^2(\mu)} = 1\}$ be the L^2 operator norm of $P_t - \mu$. Then there exists $c > 0$ such that*

$$\frac{1}{c}(1 + t^{2(N-1)})e^{-2\rho t} \leq \|P_t - \mu\|^2 \leq c(1 + t^{2(N-1)})e^{-2\rho t}$$

Proof. Since $\|P_t f - \mu f\|_{L^2(\mu)}^2 = \operatorname{Ent}_\mu^a(P_t f)$ with $a(f) = f^2$, the upper bound is given by Corollary 12. For the lower bound it is sufficient to exhibit a particular function f such that this holds. For $u \in \mathbb{R}^d$, consider the linear function $f(x) = u^T x$ and let $u_t = e^{-tB^T} u$. Then $\nabla(u_t^T x) = u_t$ and $\operatorname{div}(D \nabla(u_t^T x)) = 0$ for all $x \in \mathbb{R}^d, t \geq 0$, so that $P_t f(x) = u_t^T x$ is explicit. In particular $\mu f = \lim_{t \rightarrow \infty} P_t f(x) = 0$, and $\|(P_t - \mu)f\|^2 = \int |x^T e^{-tB^T} u|^2 \mu(dx)$. If u is a generalized eigenvector of B^T of order N associated to $\lambda \in \sigma(B^T)$ with $\Re(\lambda) = \rho$ and if $u_i = (B^T - \lambda I)^i u$ for $i \in \llbracket 0, N \rrbracket$ then $u_t = e^{-\rho t} \sum_{k=0}^{N-1} \frac{t^k}{k!} u_k$. At large time the leading term is $t^{N-1} u_{N-1}$, and so

$$\begin{aligned} \int |x^T e^{-tB^T} u|^2 \mu(dx) &\geq \int |x^T e^{-tB^T} u|^2 1_{\left\{ \frac{|x^T u_{N-1}|}{|x| |u_{N-1}|} \geq \frac{1}{2} \right\}} \mu(dx) \\ &\geq \gamma(1 + t^{2N}) e^{-2\rho t} \int |x|^2 1_{\left\{ \frac{|x^T u_{N-1}|}{|x| |u_{N-1}|} \geq \frac{1}{2} \right\}} \mu(dx) \end{aligned}$$

for some $\gamma > 0$, which concludes. \square

Remark: Note that the proof of Inequality (22) does not use the positivity of ρ , and therefore it holds for $\rho \leq 0$. Of course in this case the process is not ergodic, but we still get a local inequality

$$a(P_t f) - P_t a(f) \leq \left(\tilde{c} \int_0^t (1 + s^{2(N-1)}) e^{-2\rho s} ds \right) P_t (a''(f) |\nabla f|^2)$$

and, in the hypoelliptic case, the Wasserstein bound

$$\mathcal{W}_2^2(\nu_1 P_t, \nu_2 P_t) \leq \tilde{c}(1 + t^{2(N-1)})e^{-2\rho t} \mathcal{W}_2^2(\nu_1, \nu_2).$$

4 Interacting particles on a graph

4.1 Settings

Let $G = \{0, \dots, N\}$ be a finite non-oriented irreducible graph with edges $E \subset G^2$. We write $i \sim j$ when two vertices i and j are neighbours, namely when $(i, j) \in E$ (and so $(j, i) \in E$). For all $(i, j) \in E$, let $W_{i,j} = W_{j,i} \in \mathcal{C}^\infty(\mathbb{R}^d)$ be an even function whose all derivatives grow at most polynomially at infinity. We suppose $W_{i,j}$ is strongly convex, which means there exists $\lambda_{i,j} > 0$ such that $y^T \nabla^2 W_{i,j}(x) y \geq \lambda_{i,j} |y|^2$ for all $x, y \in \mathbb{R}^d$ where $\nabla^2 W_{i,j}$ is the Hessian matrix of $W_{i,j}$. We call $W_{i,j}$ the interaction potential between the sites i and j .

We sum up the assumptions that implicitly holds throughout Section 4:

Assumptions.

(A) *The graph G is finite, irreducible, non-oriented.*

(B) *For all $i \sim j$, $W_{i,j} = W_{j,i}$ is a smooth, even, $\lambda_{i,j}$ -strictly convex potential with $\lambda_{i,j} > 0$ whose all derivatives grow at most polynomially at infinity. For $i \not\sim j$, set $W_{i,j} = 0$.*

We are interested in $\mathbf{X}(t) = (X_i(t))_{i \in G} \in \mathbb{R}^{dN}$, a diffusion indexed by G . At time t , the particle at site i undergoes an attracting force $-\nabla W_{i,j}(X_i(t) - X_j(t))$ from the particle at site j . Moreover, depending on their site, the particles undergo infinitesimal collisions from thermal motion (see [17] for more considerations on the model). The Hamiltonian dynamics should be

$$\forall i \in G, \quad \begin{cases} dX_i(t) &= Y_i(t)dt \\ m_i dY_i(t) &= -\sum_{i \sim j} \nabla W_{i,j}(X_i(t) - X_j(t)) dt - \nu_i Y_i(t)dt + \sum_{j=1}^{\infty} \sigma_{i,j} dB_j(t) \end{cases}$$

where $Y_i(t) \in \mathbb{R}^d$ is the velocity of the particle i , $m_i > 0$ is its mass, $\nu_i > 0$ a friction coefficient, $\mathbf{B}(t) = (B_j(t))_{j \geq 1}$ is a sequence of independant standard 1-dimensional Brownian motions and for all $i \in G$, $(\sigma_{i,j})_{j \geq 1} \in \ell^2(\mathbb{R}^d)$.

Following the ideas of Arnold and Erb [2, Section 7], we could tackle this Hamiltonian process in the case where for all $(i, j) \in E$, $W_{i,j}$ is a perturbation of a quadratic potential. Nevertheless we will rather consider a simpler case without any restriction on $W_{i,j}$ other than strong convexity. Taking $\nu_i = 1$ and letting the masses m_i go to zero, we obtain the overdamped dynamics

$$\forall i \in G, \quad dX_i(t) = -\sum_{i \sim j} \nabla W_{i,j}(X_i(t) - X_j(t)) dt + \sum_{j=1}^{\infty} \sigma_{i,j} dB_j(t) \quad (23)$$

For $x = (x_1, \dots, x_N) \in \mathbb{R}^{dN}$ and $i \in G$, let $\mathbf{W}(x) = \sum_{k \in G} \sum_{j \sim k} W_{k,j}(x_k - x_j)$,

$$b_i(x) = \sum_{i \sim j} \nabla W_{i,j}(x_i - x_j) = \frac{1}{2} \nabla_i \mathbf{W}(x)$$

and $b = (b_0, \dots, b_N) = \frac{1}{2} \nabla \mathbf{W}$; let $\sigma_j = (\sigma_{0,j}, \dots, \sigma_{N,j})$ and $S = \frac{1}{2} \sum_{j \geq 1} \sigma_j \sigma_j^T$. Then the generator associated to (23) is

$$Lf(x) = -b(x) \cdot \nabla f(x) + \operatorname{div}(S \nabla f). \quad (24)$$

Let $\bar{X}(t) = \frac{1}{N+1} \sum_{i \in G} X_i(t)$. Since $W_{i,j} = W_{j,i}$ is an even function,

$$\begin{aligned} d\bar{X}(t) &= \frac{1}{N+1} \sum_{j \geq 1} \left(\sum_{i \in G} \sigma_{i,j} \right) dB_j(t) \\ \Rightarrow \quad \bar{X}(t) &= \bar{X}(0) + \frac{1}{N+1} \sum_{j \geq 1} \left(\sum_{i \in G} \sigma_{i,j} \right) B_j(t) \end{aligned}$$

If $\sum_{i \in G} \sigma_{i,j} = 0$ for all j the process \mathbf{X} is not hypoelliptic, since \bar{X}_t is constant. On the other hand if $\sum_{i \in G} \sigma_{i,j} \neq 0$ for some j , \mathbf{X} is not recurrent since \bar{X}_t is not. There are several natural ways to force the recurrence:

1. We can suppose that some particles, rather than following (23), are fixed at the origin. In this case we can always merge these particles in order to consider only X_0 is fixed. We call this *the fixed problem associated to (24)*, and we call $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)$ the associated process (more generally, if $x = (x_0, \dots, x_N) \in \mathbb{R}^{(N+1)d}$, we note $\tilde{x} = (x_1, \dots, x_N)$). The generator \tilde{L} of $\tilde{\mathbf{X}}$ is obtained from (24) by replacing the drift $b(x)$ by $\tilde{b}(0, \tilde{x})$ and the matrix $S = (s_{i,j})_{0 \leq i,j \leq N}$ by $\tilde{S} = (s_{i,j})_{1 \leq i,j \leq N}$ (where we decompose S as a $(N+1) \times (N+1)$ square of $d \times d$ blocks).
2. We can add a coercive force $-\nabla U_i(X_i(t))$ with a strongly convex U to the dynamics of some particles. In this case we can add a particle $X_{-1}(t)$ which is fixed at zero and consider that $U_i(X_i(t)) = W_{i,-1}(X_i(t) - X_{-1}(t))$, so that this case is equivalent to the previous one.
3. We can observe the cloud of particles from its center of mass, meaning that $X_i(t)$ is replaced by $X_i(t) - \bar{X}(t)$. We call this *the centered problem associated to (24)*, and $\hat{\mathbf{X}} = \mathbf{X} - \bar{X} = (X_i - \bar{X})_{i=0..N}$ (more generally, if $x \in \mathbb{R}^d$, we note $\bar{x} = \frac{1}{N+1} \sum x_i$ and $\hat{x} = x - \bar{x}$). The generator \hat{L} of the centered process is obtained from (24) by replacing the diffusion vectors σ_j by $\hat{\sigma}_j$. This process can never be hypoelliptic, and thus we may also consider $\hat{\mathbf{X}} + Z$ where Z is a standard Ornstein-Uhlenbeck process on \mathbb{R}^d satisfying $dZ = -Zdt + dB'(t)$ where B' is a standard Brownian motion on \mathbb{R}^d which is independant from \mathbf{B} . Note that the dynamics (24) is invariant by translation of the center of mass: if f is invariant by translation (i.e. $f(x_1 + h, \dots, x_N + h) = f(x_1, \dots, x_N) \forall h \in \mathbb{R}^d$) then

$$\mathbb{E}(f(\mathbf{X}(t))) = \mathbb{E}\left(f\left(\hat{\mathbf{X}}(t)\right)\right) = \mathbb{E}\left(f\left(\hat{\mathbf{X}} + Z(t)\right)\right).$$

Let \mathcal{A} be the space of \mathcal{C}^∞ functions on $\mathbb{R}^{(N+1)d}$ whose all derivatives grow at most polynomially at infinity, $\mathcal{A}_0 = \{f \in \mathcal{A}, \nabla_{x_0} f = 0\}$ which may be seen as a set of functions on \mathbb{R}^{dN} , and

$$\mathcal{A}_c = \left\{ f \in \mathcal{A}, f(x+h) = f(x) \forall x \in \mathbb{R}^{(N+1)d}, h \in \mathbb{R}^d \right\} = \left\{ f \in \mathcal{A}, \sum_0^N \nabla_{x_i} f = 0 \right\}$$

Proposition 14. *For any initial law, the process \mathbf{X} (resp. $\tilde{\mathbf{X}}$) is well-defined for all times. The associated semi-group $P_t f(x) = \mathbb{E}(f(\mathbf{X}(t)) \mid \mathbf{X}(0) = x)$ (resp. \tilde{P}_t) is Feller and fix A_c (resp. A_0).*

Proof. The convexity and growth conditions on \mathbf{W} ensures that $\phi(x) = 1 + \mathbf{W}(x) + (\bar{x})^2$ satisfies $L\phi \leq A \leq A\phi$ for some $A > 0$, which implies equation (23) admits a unique strong solution \mathbf{X} (see [16, Theorem 3.1] ; the same holds for the fixed problem with $\phi = 1 + \tilde{\mathbf{W}}$) and the associated semi-group is Feller (see [28, (22.5) p. 164]). If f is \mathcal{C}^∞ , so is $P_t f$ for all $t \geq 0$ (see [11, Theorem VII.5]). The existence of the Lyapunov function ϕ implies the moments of \mathbf{X} grow at most linearly in time (the proof of [29, Lemma 2.1] may be adapted to our case), which means if f grows at most polynomially at infinity, so is $P_t f$ for $t \geq 0$. If n is a multi-index, differentiating with respect to the space variables the Kolmogorov equation $\partial_t P_t f = L P_t f$ yields, from [11, Theorem VII.10],

$$\partial_x^n P_t f = P_t \partial_x^n f + \int_0^t P_{t-s} [\partial^n, L] P_s f ds.$$

For some $C, k > 0$ depending on the derivatives of W , $|[\partial^n, L]g|(x) \leq C(1+x^k) \sum_{|j| < |n|} |\partial^j g|(x)$. Therefore by induction on $|n|$, if $\partial_x^j f$ grow at most polynomially at infinity for all $|j| < |n|$, so does $\partial_x^n P_t f$. The case of \tilde{P}_t is similar. \square

4.2 Main results

We note

$$L_G h(i) = \sum_{j \sim i} \lambda_{ij} (h(j) - h(i))$$

the discrete Laplacian on G (also seen as an $(N+1) \times (N+1)$ matrix). This is the generator of an ergodic Markov chain on G (recall the λ_{ij} 's are positive and G is irreducible). Since L_G is symmetric the invariant measure is the uniform law ν (also seen as the column vector $\frac{1}{N+1}(1, \dots, 1)$) on G . Let

$$\rho = \inf_{h \in \mathbb{R}^{N+1}} \frac{-h^T L_G h}{|h - \nu h|^2}$$

be the spectral gap of L_G and

$$\rho_D = \inf_{h \in \mathbb{R}^{N+1}, h_0=0} \frac{-h^T L_G h}{|h|^2}$$

be its Dirichlet eigenvalue.

Proposition 15. *For the fixed problem associated to (24), for all admissible function a and for all $t \geq 0, f \in \mathcal{A}_0$,*

$$a''(\tilde{P}_t f) |\nabla \tilde{P}_t f|^2 \leq e^{-2\rho_D t} \tilde{P}_t (a''(f) |\nabla f|^2).$$

For the centered problem associated to (24), for all $t \geq 0, f \in \mathcal{A}_c$,

$$a''(P_t f) |\nabla P_t f|^2 \leq e^{-2\rho t} P_t (a''(f) |\nabla f|^2)$$

Proof. For $\Phi(f) = a''(f)|\nabla f|^2$, we computed in (17) that

$$\Gamma_{\hat{L},\Phi}f \geq a''(f)(\nabla f)^T J_b \nabla f \quad \text{and} \quad \Gamma_{\tilde{L},\Phi}f \geq a''(f)(\nabla f)^T J_{\tilde{b}} \nabla f$$

where J_b and $J_{\tilde{b}}$ are the Jacobian matrices of the drifts b and \tilde{b} . We compute

$$\begin{aligned} (\nabla f)^T J_b \nabla f &= \sum_{i \in G} \sum_{j \sim i} (\nabla_{x_i} f)^T (\nabla^2 W_{i,j}) (\nabla_{x_i} - \nabla_{x_j}) f \\ &= \sum_{(i,j) \in E} ((\nabla_{x_i} - \nabla_{x_j}) f)^T (\nabla^2 W_{i,j}) (\nabla_{x_i} - \nabla_{x_j}) f \\ &\geq \sum_{(i,j) \in E} \lambda_{i,j} |(\nabla_{x_i} - \nabla_{x_j}) f|^2 \\ &= -(\nabla f)^T L_G \nabla f, \end{aligned}$$

and similarly $(\nabla f)^T J_{\tilde{b}} \nabla f \geq -(\nabla f)^T L_G \nabla f$ for $f \in \mathcal{A}_0$. When $f \in \mathcal{A}_c$, $\nu \nabla f = 0$, and thus $\Gamma_{\hat{L},\Phi}f \geq \rho \Phi(f)$. When $f \in \mathcal{A}_0$, $\Gamma_{\tilde{L},\Phi}f \geq \rho_D \Phi(f)$. In both case (and since \mathcal{A}_c and \mathcal{A}_0 are respectively fixed by P_t and \tilde{P}_t), Lemma 1 concludes. \square

Let

$$\Theta_N = \{x \in \mathbb{R}^{(N+1)d}, \bar{x} = 0\}$$

so that $\hat{\mathbf{X}} \in \Theta_N$ for all $t \geq 0$, and thus the associated semi-group \hat{P}_t acts on $\mathcal{P}(\Theta_N)$.

Corollary 16. *Suppose in addition to Assumptions (A) and (B) that the processes \mathbf{X} and $\tilde{\mathbf{X}}$ are hypoelliptic. Then for any initial laws $\nu_1, \nu_2 \in \mathcal{P}(\Theta_N)$,*

$$\mathcal{W}_2(\nu_1 \hat{P}_t, \nu_2 \hat{P}_t) \leq e^{-\rho t} \mathcal{W}_2(\nu_1, \nu_2)$$

and for any initial laws $\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^{dN})$,

$$\mathcal{W}_2(\nu_1 \tilde{P}_t, \nu_2 \tilde{P}_t) \leq e^{-\rho_D t} \mathcal{W}_2(\nu_1, \nu_2).$$

Moreover both the fixed and the centered problems admits a unique invariant law, respectively denoted by μ_0 and μ_c , which satisfy the log-Sobolev inequalities

$$\begin{aligned} \forall f \in \mathcal{A}_c \text{ s.t. } \int f d\mu_c = 1, \quad \int f \ln f d\mu_c &\leq \frac{|S|}{2\rho} \int \frac{|\nabla f|^2}{f} d\mu_c \\ \forall f \in \mathcal{A}_0 \text{ s.t. } \int f d\mu_0 = 1, \quad \int f \ln f d\mu_0 &\leq \frac{|\tilde{S}|}{2\rho_D} \int \frac{|\nabla f|^2}{f} d\mu_0 \end{aligned}$$

Proof. Both cases are similar and thus we only consider the first one. The Wasserstein contractions is implied by Theorem 6 and Proposition 15 for $a(f) = f^2$ (recall that for $f \in \mathcal{A}_c$, $P_t f = \hat{P}_t f$). When E is a Polish space, so is $\mathcal{P}_2(E) = \{\nu \in \mathcal{P}(E) \text{ having a finite } 2^{nd} \text{ moment}\}$ endowed with the \mathcal{W}_2 metric ([8]). Therefore by the fixed point Theorem, $\nu = \nu \hat{P}_t$ admits a unique solution in $\mathcal{P}_2(\Theta_N)$ which may a priori depend on t , but then $\nu \hat{P}_s = \nu \hat{P}_t \hat{P}_s = \nu \hat{P}_s \hat{P}_t$, and by uniqueness of the solution $\nu \hat{P}_s = \nu$ for all $s \geq 0$. The existence of the Lyapunov function \mathbf{W} implies an invariant measure necessarily have a finite second moment.

The log-Sobolev inequality is implied by Proposition 15 for $a(x) = x \ln x$ together with Lemma 2 for $\gamma(t) = e^{-2\rho t}$, $\phi_1(f) = f \ln f$ and $\phi_2(f) = \frac{|\nabla f|^2}{f}$ so that

$$\Gamma_{\Phi_1} f = \frac{(\nabla f)^T S \nabla f}{2f} \leq \frac{1}{2} |S| \Phi_2(f).$$

\square

Remark: since the drift of the diffusion is of the form $-\nabla \mathbf{W}$ with a convex \mathbf{W} , there would be no difficulties to apply here the Meyn-Tweedie-Lyapunov techniques for proving convergence in total variation distance (cf. [13, 3]) under controllability or strong hypoellipticity conditions. However the "explicit" speed of convergence obtained this way depend on estimates of the probability transition of the process on a compact set, which are much less tracktable than the constant involved in Corollary 16. The use of an explicit mirror coupling (see [12]) may give satisfactory estimates.

4.3 Chain of particles

Suppose $i \sim j$ iff $|i - j| \leq 1$, and $W_{i,j} = W$ does not depend on the edge (i, j) (and so does $\lambda_{i,j} = \lambda$). We call this dynamics *the chain of $N + 1$ particles*.

Proposition 17. *For the chain of $N + 1$ particles, for $N \geq 3$,*

$$\rho \geq \frac{\lambda}{(N + 1)^2}$$

Proof. Let $h = \min \left\{ \frac{|\partial A|}{|A|}, |A| < \frac{|G|}{2} \right\}$ be the Cheeger constant of G , where $\partial A \subset E$ is the set of (non-oriented) edges (i, j) with $i \in A$ and $j \notin A$, and $|B|$ denotes the cardinal of a set B . Since the maximal degree of G is 2, from [25, Theorem 4.2], for $N \geq 3$,

$$\rho \geq \lambda \frac{h^2}{4}.$$

It is clear the minimum of $\frac{|\partial A|}{|A|}$ is attained with $A = \{0, \dots, \lfloor (N+1)/2 \rfloor\}$, so that $h \geq \frac{2}{N+1}$. \square

Proposition 18. *For the chain of $N + 1$ particles,*

$$\rho_D \geq \lambda \left(1 - \cos \left(\frac{\pi}{2N} \right) \right)$$

Proof. Consider $(Y_t)_{t \geq 0}$ the continuous time nearest neighbour random walk on $\{0, \dots, N\}$ starting at N and absorbed at 0, whose generator is $\frac{1}{\lambda} \tilde{L}_G$. The dynamics is the following: from the time t_0 , Y waits a time E with standard exponential law, and then jump to Y_{t_0+E} equal to either $Y_{t_0} + 1$ or $Y_{t_0} - 1$ with equal probability $\frac{1}{2}$, unless $Y_{t_0} = 0$ in which case it does not jump, or unless $Y_{t_0} = N$ in which case it jumps to $N - 1$ with probability 1. According to the work [14] of Diaconis and Miclo, if T is the time of absorption of Y , then

$$\lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{P}(T > t) = -\rho_D.$$

The process Y is very close to the process \tilde{Y} of [14, Example 16], the only difference (up to renumbering the states) being that when $\tilde{Y}_{t_0} = N$, the process \tilde{Y} only waits a time $\frac{1}{2}E$ to bounce back to $N - 1$. As a consequence, \tilde{Y} goes faster to N than Y , but $(Y_{2t})_{t \geq 0}$ goes faster to N than \tilde{Y}_t and

$$\lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{P}(T > t) \leq \frac{1}{2} \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{P}(\tilde{T} > t)$$

where \tilde{T} is the absorbing time of \tilde{Y} . Now explicit spectral computations (cf. [14]) yields

$$\lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{P}(\tilde{T} > t) = -2 \left(1 - \cos \left(\frac{\pi}{2N} \right) \right).$$

\square

Remark: As an alternative proof, we could have estimated the Laplace transform of T . Indeed Chernoff's inequality yields $\mathbb{P}(T > t) \leq \mathbb{E}(e^{\theta T}) e^{-\theta t}$. Denoting by T_k the absorbing time starting from k and $b_k = \mathbb{E}(e^{\theta T_k}) (\mathbb{E}(e^{\theta T_N}))^{-1}$, a recurrence relation satisfied by the b_k 's and the fact $T_0 = 0$ allow to prove that for any $\eta < \frac{\pi}{2N}$, $\mathbb{E}(e^{(1-\cos(\eta))\lambda T_k}) = \frac{1}{\cos(k\eta)}$. This means $\rho_D \geq \lambda(1 - \cos(\frac{\pi}{2N}))$.

Now consider a chain of particles for which all the diffusion blocks $s_{i,j}$ are zero except $s_{N,N}$ and $s_{0,0}$ which are homogeneous dilations. In other words, we are interested in the processes \mathbf{X} and $\tilde{\mathbf{X}}$ defined by

$$(25) \quad \left\{ \begin{array}{ll} dZ_i &= -(\nabla W(Z_i - Z_{i-1}) + \nabla W(Z_i - Z_{i+1})) dt, \quad \forall i \in \llbracket 1, N-1 \rrbracket, \quad Z_i = X_i \text{ or } \tilde{X}_i \\ dZ_N &= -\nabla W(Z_N - Z_{N-1}) dt + \sigma_N dB_1(t), \quad Z_N = X_N \text{ or } \tilde{X}_N \\ dX_0 &= -\nabla W(X_0 - X_i) dt + \sigma_0 dB_2(t) \\ \tilde{X}_0 &= 0. \end{array} \right.$$

where B_1 and B_2 are independant d -dimensional Brownian motion and $\sigma_0, \sigma_N \in \mathbb{R}$.

Lemma 19. *The processes \mathbf{X} and $\tilde{\mathbf{X}}$ defined by (25) are hypoelliptic. Moreover, if $\nabla^2 W$ is bounded, then the assumptions of Theorem 9 are satisfied.*

Proof. Again both cases are similar and we only treat the centered problem. Let $V_0 = \sigma_N^2 \nabla_N$ and for $i \in \llbracket 1, N \rrbracket$ let $V_i = \sigma_N^2 \nabla^2 W(x_{N-1} - x_N) \dots \nabla^2 W(x_{N-1-i} - x_{N-i}) \nabla_{N-i}$. Then $[V_i, \nabla \mathbf{W} \cdot \nabla] = V_{i+1} + r_i$ where $r_i(x) \in \text{span}\{V_j(x), j \leq i\}$ for all $x \in \mathbb{R}^{d(N+1)}$ and the parabolic Hörmander's condition is fulfilled for \mathbf{X} .

If moreover $\nabla^2 W$ is bounded, from

$$[-\nabla \mathbf{W} \cdot \nabla, \nabla_j] = \nabla^2 W(x_j - x_{j-1}) (\nabla_{j-1} - \nabla_j) + \nabla^2 W(x_{j+1} - x_j) (\nabla_{j+1} - \nabla_j),$$

the assumptions of Theorem 9 hold with $C_i = \nabla_{N-i}$ and $m = 2\Lambda = 2\|\nabla^2 W\|_\infty$. \square

As a conclusion, Corollary 16 may be applied to \mathbf{X} and $\tilde{\mathbf{X}}$ with $|S| = \max(\sigma_N^2, \sigma_0^2)$, $|\tilde{S}| = \sigma_N^2$, and with ρ and ρ_D bounded in Propositions 17 and 18. If moreover $\nabla^2 W$ is bounded, both Theorems 9 and 10 holds and we get for $\mathbf{X} - \tilde{X}$ and $\tilde{\mathbf{X}}$ convergences of the form

$$\text{Ent} P_t f \leq c N^{c' N^2} e^{-\frac{c''}{N^2} t} \text{Ent} f$$

where c, c' and c'' do not depend on N .

References

- [1] F. Achleitner, A. Arnold, and D. Stürzer. Large-time behavior in non-symmetric Fokker-Planck equations. *ArXiv e-prints*, June 2015.
- [2] A. Arnold and J. Erb. Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift. *ArXiv e-prints*, September 2014.

- [3] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.
- [4] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.
- [5] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [6] J.-B. Bardet, A. Christen, A. Guillin, F. Malrieu, and P.-A. Zitt. Total variation estimates for the TCP process. *Electron. J. Probab.*, 18:no. 10, 21, 2013.
- [7] F. Baudoin. Bakry-Emery meet Villani. *ArXiv e-prints*, August 2013.
- [8] F. Bolley. Separability and completeness for the Wasserstein distance. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 371–377. Springer, Berlin, 2008.
- [9] F. Bolley and I. Gentil. Phi-entropy inequalities for diffusion semigroups. *J. Math. Pures Appl. (9)*, 93(5):449–473, 2010.
- [10] E. Bouin, F. Hoffmann, and C. Mouhot. Rate of convergence to equilibrium for a model describing fibre lay-down processes on a moving conveyor belt. (*on-going work*), 2015.
- [11] H. Brezis. *Analyse fonctionnelle. Théorie et applications*. Collection Mathématiques Appliquées pour la Maîtrise.. [Collection of Applied Mathematics for the Master’s Degree]. Masson, Paris, 1983.
- [12] P. Cattiaux and A. Guillin. Semi log-concave Markov diffusions. In *Séminaire de Probabilités XLVI*, volume 2123 of *Lecture Notes in Math.*, pages 231–292. Springer, Cham, 2014.
- [13] P. Cattiaux, A. Guillin, and P.-A. Zitt. Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1):95–118, 2013.
- [14] P. Diaconis and L. Miclo. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.*, 22(3):558–586, 2009.
- [15] J. Dolbeault, C. Mouhot, and C. Schmeiser. Hypocoercivity for kinetic equations with linear relaxation terms. *C. R. Math. Acad. Sci. Paris*, 347(9-10):511–516, 2009.
- [16] R. Durrett. *Stochastic calculus. A practical introduction*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996.
- [17] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.
- [18] S. Gadat and L. Miclo. Spectral decompositions and \mathbb{L}^2 -operator norms of toy hypocoercive semi-groups. *Kinet. Relat. Models*, 6(2):317–372, 2013.
- [19] M. Hairer. How hot can a heat bath get? *Comm. Math. Phys.*, 292(1):131–177, 2009.

- [20] F. Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications. *J. Funct. Anal.*, 244(1):95–118, 2007.
- [21] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.
- [22] K. Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.*, 258(11):3758–3774, 2010.
- [23] K. Kuwada. Gradient estimate for Markov kernels, Wasserstein control and Hopf-Lax formula. In *Potential theory and its related fields*, RIMS Kôkyûroku Bessatsu, B43, pages 61–80. Res. Inst. Math. Sci. (RIMS), Kyoto, 2013.
- [24] F. Malrieu. Some simple but challenging markov processes. *preprint*.
- [25] B. Mohar. Isoperimetric numbers of graphs. *J. Combin. Theory Ser. B*, 47(3):274–291, 1989.
- [26] P. Monmarché. On \mathcal{H}^1 and entropic convergence for contractive PDMP. *ArXiv e-prints*, April 2014.
- [27] P. Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *ArXiv e-prints*, February 2015.
- [28] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2 Itô calculus*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Reprint of the second (1994) edition.
- [29] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.
- [30] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.